

A Novel Analytical Method for Evolutionary Graph Theory Problems

Paulo Shakarian

*Network Science Center and Dept. of Electrical Engineering and Computer Science,
United States Military Academy, West Point, NY 10996*

Patrick Roos

Dept. of Computer Science, University of Maryland, College Park, MD 20740

Geoffrey Moores

*Network Science Center and Dept. of Electrical Engineering and Computer Science,
United States Military Academy, West Point, NY 10996*

Abstract

Evolutionary graph theory studies the evolutionary dynamics of populations structured on graphs. A central problem is determining the probability that a small number of mutants overtake a population. Currently, Monte Carlo simulations are used for estimating such fixation probabilities on general directed graphs, since no good analytical methods exist. In this paper, we introduce a novel deterministic framework for computing fixation probabilities for strongly connected, directed, weighted evolutionary graphs under neutral drift. We show how this framework can also be used to calculate the expected number of mutants at a given time step (even if we relax the assumption that the graph is strongly connected), how it can extend to other related models (e.g. voter model), how our framework can provide non-trivial bounds for fixation probability in the case of an advantageous mutant, and how it can be used to find a non-trivial lower bound on the mean time to fixation. We provide various experimental results determining fixation probabilities and expected number of mutants on different graphs. Among these, we show that our method consistently outperforms Monte Carlo simulations in speed by several orders of magnitude. Finally we show how our approach can provide insight into synaptic competition in neurology.

Keywords: evolutionary dynamics, Moran process, complex networks

1. Introduction

Evolutionary graph theory (EGT), introduced by [14], studies the problems related to population dynamics when the underlying structure of the population is represented as a directed, weighted graph. This model has been applied to problems in evolutionary biology [28], physics [25], game theory [20], neurology [26], and distributes systems [13]. A central problem in this research area is computing the *fixation probability* - the probability that a certain subset of mutants overtakes the population. Although good analytical approximations are available for the undirected/unweighted case [1, 6], these break down for directed, weighted graphs as shown by [16]. As a result, most work dealing with evolutionary graphs rely on Monte Carlo simulations to approximate the fixation probability [22, 7, 4]. In this paper we develop a novel deterministic framework to compute fixation probability in the case of neutral drift (when mutants and residents have equal fitness) in directed, weighted evolutionary graphs based on the convergence of “vertex probabilities” to the fixation probability as time approaches infinity. We then show how this framework can be used to calculate the expected number of mutants at a given time, how the framework can be modified to do the same for related models, how it can provide non-trivial bounds for fixation probability in the case of an advantageous mutant, and how it can provide a non-trivial lower bound on the mean time to fixation. We also provide various experiments that show how our method can outperform Monte Carlo simulations by several orders of magnitude. Additionally, we show that the results of this paper can provide direct insight into the problem of synaptic competition in neurology.

Our method also fills a few holes in the literature. First, it allows for deterministic computation of fixation probability when there is an initial *set* of mutants – not just a singleton (the majority of current research on evolutionary graph theory only considers singletons). Second, it allows us to study how the mutant population changes as a function of time. Third, we show (by way of rigorous proof) that fixation probability, under the case of neutral drift is a lower bound for the case of the advantageous mutant – confirming simulation observations by [15]. Fourth, we show (also by way of rigorous proof) that fixation probability under neutral drift is additive (even

for weighted, directed graphs), which extends the work of [6] that proved this for undirected/unweighted graphs. Fifth, we provide a non-trivial lower bound for the computation of mean time to fixation in the general case - which has only previously been explored for well-mixed populations [2] and special cases of graphs [5].

This paper is organized as follows. In Section 2 we review the original model of Lieberman et al., introduce the idea of “vertex probabilities” and show how they can be used to find the fixation probability. We then show how this can be used to determine the expected number of mutants at a given time in Section 3. This is followed by a discussion of how the framework can be extended to other update rules in Section 4 and then for bounding fixation probability in the case of an advantageous mutant in Section 5. We then discuss how our approach can be adopted to bound mean time to fixation in Section 6. We use the results of the previous sections to create an algorithm for computing fixation probability and introduce a heuristic technique to significantly decrease the run-time. The algorithm and several experimental evaluations are described in Section 7. In Section 8, we show how our framework can be applied to neurology to gain insights into synaptic competition. Finally, we discuss related work in Section 9 and conclude.

2. Directly Calculating Fixation Probability

The classic evolutionary model known as the *Moran Process* is a stochastic process used to describe evolution in a well-mixed population [18]. All the individuals in the population are either *mutants* or *residents*. The aim of such work was to determine if a set of mutants could take over a population of residents (achieving “fixation”). In [14], evolutionary graph theory (EGT) is introduced, which generalizes the model of the Moran Process by specifying relationships between the N individuals of the population in the form of a directed, weighted graph. Here, the graph will be specified in the usual way as $G = (V, E)$ where V is a set of nodes (individuals) and $E \subseteq V \times V$. In most literature on evolutionary graph theory, the evolutionary graph is assumed to be *strongly connected*. We make the same assumption and state when it can be relaxed.

For any node i , the numbers $k_{in}^{(i)}, k_{out}^{(i)}$ are the in- and out- degrees respectively. We will use the symbol N to denote the size of V . Additionally, we will specify weights on the edges in set E using a square matrix denoted $W = [w_{ij}]$ whose side is of size N . Intuitively, w_{ij} is the probability that

member of the population j is replaced by i given that member i is selected. We require $\sum_j w_{ij} = 1$ and that $(i, j) \in E$ iff $w_{ij} > 0$. If for all i, j , we have $w_{ij} = 1/k_{out}^{(i)}$, then the graph is said to be “unweighted.” If for all $(i, j) \in E$, we have $(j, i) \in E$ the graph is said to be “undirected.” Though our results primarily focus on the general case, we will often refer to the special case of undirected/unweighted graphs as this special case is quite common in the literature [1, 6].

In this paper we will often consider the outcome of the evolutionary process when there is a set of initial mutants as opposed to a singleton. Hence, we say some set (often denoted C) is a *configuration* if that set specifies the set of mutants in the population (all other members in the population then are residents). We assume all members in the population are either mutants or residents and have a fitness specified by a parameter $r > 0$. Mutants have a fitness r and residents have a fitness of 1. At each time step, some individual $i \in V$ is selected for “birth” with a probability proportional to its fitness. Then, an outgoing neighbor j of i is selected with probability w_{ij} and replaced by a clone of i . Note if $r = 1$, we say we are in the special case of *neutral drift*.

We will use the notation $P_{V',t}$ to refer to the probability of being in configuration V' after t timesteps and $P_{V',t|C}$ to be the probability of being in configuration V' at time t conditioned upon initial configuration C . Perhaps the most widely studied problem in evolutionary graph theory is to determine the fixation probability. Given set of mutants C at time 0, the fixation probability is defined as follows.

$$F_C = \lim_{t \rightarrow \infty} P_{V',t|C} \quad (1)$$

This is the probability that an initial set C of mutants takes over the entire population as time approaches infinity. Similarly, we will use the term the *extinction probability*, \overline{F}_C , to be $\lim_{t \rightarrow \infty} P_{\emptyset,t|C}$. If the graph is strongly connected, then we have $F_C + \overline{F}_C = 1$. Hence, for a strongly connected graph, a mutant either fixates or becomes extinct. Typically, this problem is studied using Monte Carlo simulation. This work uses the idea of a *vertex probabilities* to create an alternative to such an approach. The vertex probability is the probability that a certain vertex is a mutant at a certain time given an initial configuration. For vertex i at time t , we denote this as $P_{i,t|C}$. Often, for ease of notation, we shall assume that the probabilities are conditioned on some initial configuration and drop the condition, writing $P_{i,t}$

instead of $P_{i,t|C}$. We note that $P_{i,t}$ can be expressed in terms of probabilities of configurations as follows.

$$P_{i,t} = \sum_{\substack{V' \in 2^V \\ s.t. \ i \in V'}} P_{V',t} \quad (2)$$

Viewing the probability that a specific vertex is a mutant at a given time has not, to our knowledge, been studied before with respect to evolutionary graph theory (or in related processes such as the voter model). The key insight of this paper is that studying these probabilities sheds new light on the problem of calculating fixation probabilities in addition to providing other insights into EGT. For example, it is easy to show the following relationship.

Proposition 1. *Let V' be a subset of V and t be an arbitrary time point. Iff for all $i \in V'$, $P_{i,t} = 1$ and for all $i \notin V'$, $P_{i,t} = 0$, then $P_{V',t} = 1$ and for all $V'' \in 2^V$ s.t. $V'' \not\equiv V'$, $P_{V'',t} = 0$.*

It is easy to verify that $F_C > 0$ iff $\forall i \in V$, $\lim_{t \rightarrow \infty} P_{i,t} > 0$. Hence, in this paper, we shall generally assume that $\lim_{t \rightarrow \infty} P_{i,t} > 0$ holds for all vertices i and specifically state when it does not. As an aside, for a given graph, this assumption can be easily checked: simply ensure for $j \in V - C$ that exists some $i \in C$ s.t. there is a directed path from i to j .

Now that we have introduced the model and the idea of vertex probabilities we will show how to leverage this information to compute fixation probability. It is easy to show that as time approaches infinity, the vertex probabilities for all vertices converge to the fixation probability when the graph is strongly connected.

Theorem 1. $\forall i, \lim_{t \rightarrow \infty} P_{i,t|C} = F_C$

Now let us consider how to calculate $P_{i,t}$ for some i and t . For $t = 0$, where we know that we are in the state where only vertices in a given set are mutants, we need only appeal to Proposition 1 - which tells us that we assign a probability of 1 to all elements in that set and 0 otherwise. For subsequent timesteps, we have developed Theorem 2 shown next (the proof of which is included in the supplement).

Theorem 2.

$$P_{i,t} = P_{i,t-1} + \sum_{(j,i) \in E} w_{ji} (P_{j,t-1} \cdot S_{(j,t)|(i,t-1)} - P_{i,t-1} \cdot S_{(j,t)|(i,t-1)})$$

($S_{(j,t)|(i,t-1)}$ is the probability that j is picked for reproduction at time t given that i was a mutant at time $t-1$.)

We believe that a concise, tractable analytical solution for $S_{(j,t)|(i,t-1)}$ is unlikely. However, for neutral drift ($r = 1$), these conditional probabilities are trivial - specifically, we have for all i, j, t , $S_{(j,t)|(i,t-1)} = 1/N$ as this probability of selection is independent of the current set of mutants or residents in the graph. Hence, in the case of neutral drift, we have the following:

$$P_{i,t} = P_{i,t-1} + \sum_{(j,i) \in E} \frac{w_{ji}}{N} \cdot (P_{j,t-1} - P_{i,t-1}) \quad (3)$$

Studying evolutionary graph theory under neutral drift was a central theme in several papers on EGT in the past few years [6, 15] as it provides an intuition on the effects of network topology on mutant spread. In Section 5 we examine the case of the advantageous mutant ($r > 1$). Neutral drift allows us to strengthen the statement of Equation 1 to a necessary and sufficient condition - showing that when the probabilities of all nodes are equal, then we can determine the fixation probability.

Theorem 3. *Assuming neutral drift ($r = 1$), given initial configuration C with fixation probability F_C , if at time t the quantities $P_{i,t|C}$ are equal (for all $i \in V$), then they also equal F_C .*

Therefore, under neutral drift, we can determine fixation probability when Equation 3 causes all $P_{i,t}$'s to be equal. We can also use Equation 3 to find bounds on the fixation probability for some time t by the following result that holds for any time t under neutral drift.

$$\min_i P_{i,t} \leq F_C \leq \max_i P_{i,t} \quad (4)$$

Under neutral drift, we can show that fixation probability is additive for disjoint sets. Broom et al. proved a similar result the a special case of undirected/unweighted evolutionary graphs [6]. However, our proof (contained in

the supplement) differs from theirs in that we leverage Equation 3. Further, unlike the result of Broom et al., our result applies to the more general case of weighted, directed graphs.

Theorem 4. *When $r = 1$ for disjoint sets $C, D \subseteq V$, $F_C + F_D = F_{C \cup D}$.*

3. Calculating the Expected Number of Mutants

In addition to allowing for the calculation of fixation probability, our framework can also be used to observe how the expected number of mutants changes over time. We will use the notation $\mathbf{Ex}_C^{(t)}$ to denote the expected number of mutants at time t given initial set C . Formally, this is defined below.

$$\mathbf{Ex}_C^{(t)} = \sum_{i \in V} P_{i,t} \quad (5)$$

Unlike fixation probability, which only considers the probability that mutants overtake a population, $\mathbf{Ex}_C^{(t)}$ provides a probabilistic average of the number of mutants in the population under a finite time horizon. For example, it has been noted that graph structures which amplify fixation normally also increase time to absorption [8, 21]. Hence, finding the expected number of mutants may be a more viable topic in some areas of research where time is known to be limited. Following from Equation 3 where we showed how to compute $P_{i,t}$ for each node at a given time, we have the following relationship concerning the expected number of mutants at a given time under neutral drift.

$$\mathbf{Ex}_C^{(t)} = \mathbf{Ex}_C^{(t-1)} + \frac{\mathbf{Ex}_C^{(t-1)}}{N} - \frac{1}{N} \sum_{i \in V} \sum_{(j,i) \in E} w_{ji} \cdot P_{i,t-1} \quad (6)$$

Based on Equation 6, we notice that for $r = 1$, at each time-step, the number of expected mutants increases by at most the average fixation probability and decreases by a quantity related to the average “temperature.” The temperature of vertex i (denoted T_i) is defined for a given node is the sum of the incoming edge weights [14]: $T_i = \sum_j w_{ji}$. Intuitively, nodes with a higher temperature change more often between being a mutant and being a

resident than those with lower temperature. Re-writing Equation 6 in terms of temperature we have the following:

$$\mathbf{Ex}_C^{(t)} = \mathbf{Ex}_C^{(t-1)} + \frac{\mathbf{Ex}_C^{(t-1)}}{N} - \frac{1}{N} \sum_{i \in V} T_i \cdot P_{i,t-1} \quad (7)$$

Hence, if the preponderance of high temperature nodes are likely to be mutants, then most likely the average number of mutants will decrease at the next time step. We also note that Theorem 2, Equation 3, and Equation 6 do not depend on the assumption that the underlying graph is strongly connected. Therefore, as such is the case, we can study the relationship of time vs. expected number of mutants for *any* evolutionary graph (under neutral drift). This could be of particular interest to non-strongly connected evolutionary graphs that may have trivial fixation probabilities (i.e. 1 or 0) but may have varying levels of mutants before achieving an absorbing state.

4. Applying the Framework to Other Update Rules

The results of the last two sections not only apply to the original model of [14], but several other related models in the literature. Viewing an evolutionary graph problem as a stochastic process, where the states represent different mutant-resident configurations, it is apparent that the original model specifies the transition probabilities. However, there are other ways to specify the transition probabilities known as update rules. Several works address different update rules [1, 25, 15]. Overall, we have identified three major families of update rules - birth-death (a.k.a. the invasion process) where the node to reproduce is chosen first, death-birth (a.k.a. the voter model) where the node to die is chosen first, and link dynamics, where an edge is chosen. We summarize these in Table 1.

We have already shown how our methods can deal with the original model of Lieberman et al., often referred to as the Birth-Death (BD) process. In this section, we apply our methods to the neutral-drift (non-biased) cases of death-birth and link-dynamics. In these models, the weights of the edges is typically not considered. Hence, in order to align this work with the majority of literature on those models, we will express vertex probabilities in terms of node in-degree ($k_{in}^{(i)}$) and the set of directed edges (E). We note that these results can be easily extended to a more general case with an edge-weight matrix as we used for the original model of EGT.

Table 1: Different families of update rules.

Update Rule	Intuition
Birth-Death (BD) (a.k.a. Invasion Process (IP))	(1) Node i selected (2) neighbor of i , node j selected (3) Offspring of i replaces j
Death-Birth (DB) (a.k.a. Voter Model (VM))	(1) Node i selected (2) neighbor of i , node j selected (3) Offspring of j replaces i
Link Dynamics (LD)	(1) Edge (i, j) selected (2) The offspring of one node in the edge replaces the other node

4.1. Death-Birth Updating

Under the death birth model (DB), at each time step, a vertex i is selected for death. With a death-bias (DB-D), it is selected proportional to the inverse of its fitness, with a birth-bias (DB-B) it is selected with a probability $1/N$, which is also the probability under neutral drift. Then, an incoming neighbor (j) is selected either proportional to the fitness of all incoming neighbors (birth-bias), or with a uniform probability (in the case death-bias or neutral drift). The selected neighbor then replaces i . Here, we compute $P_{i,t}$ under this dynamic with $r = 1$.

$$P_{i,t} = (1 - N^{-1})P_{i,t-1} + (Nk_{in}^{(i)})^{-1} \sum_{(j,i) \in E} P_{j,t-1} \quad (8)$$

We note that the proof of convergence still holds for death-birth - that is for some time t , $\forall i$, the value $P_{i,t}$ is the same, then $P_{i,t} = F_C$. Further, Theorem 4 holds for DB under neutral drift as well, specifically, for disjoint sets $C, D \subseteq V$, $F_C + P_D = P_{C \cup D}$.

4.2. Link-Dynamics

With link dynamics (LD), at each time step an edge (i, j) is selected either proportional to the fitness of i or the inverse of the fitness of j . It has previously been shown that LD under birth bias is an equivalent process

to LD with a death bias [15]. Under neutral drift, the probability of edge selection is $1/|E|$ (where $|E|$ is the cardinality of set E). Then, i replaces j . Now, we compute $P_{i,t}$ under this dynamic with $r = 1$.

$$P_{i,t} = (1 - k_{in}^{(i)}|E|^{-1})P_{i,t-1} + \frac{1}{|E|} \sum_{(j,i) \in E} P_{j,t-1} \quad (9)$$

Again, convergence and additivity of the fixation probability still hold under link dynamics just as with BD and DB.

5. Bounding Fixation Probability for $r > 1$

So far we have shown how our method can be used to find fixation probabilities under the case of neutral drift. Here, we show how our framework can be useful in the case of an advantageous mutant (when the value for r , the relative fitness, is greater than 1). First, we show that our method provides a lower bound. We then provide an upper bound on the fixation probability that can be used in conjunction with our framework when studying the case of the advantageous mutant. We note that certain parts of these proofs are specific for different update rules, and we identify them using the abbreviations from the last section (DB-D, DB-B, and LD). The update of the original model of [14] is known as the “birth-death” model and abbreviated BD. If the fitness bias is on a birth event, we denote it as BD-B and if the bias is on a death event we denote it as BD-D.

Naoki Masuda observes experimentally (through simulation) that the fixation probability computed with neutral drift appears to be a lower bound on the fixation probability for an advantageous mutant [15]. We were able to prove this result analytically – the proof is included in the supplementary materials.

Theorem 5. *For a given set C , let $F_C^{(1)}$ be the fixation probability under neutral drift and $F_C^{(r)}$ be the fixation probability calculated using a mutant fitness $r > 1$. Then, under BD-B, BD-D, DB-B, DB-D, or LD dynamics, $F_C^{(1)} \leq F_C^{(r)}$.*

This proof leads to the conjecture that $r' > r$ implies $F_C^{(r')} \geq F_C^{(r)}$. However, we suspect that proving this monotonicity property will require a different technique than used in Theorem 5. Next, to find an upper bound

that corresponds with the lower bound above, we use the proof technique introduced in [9], to obtain the following non-trivial upper bounds of fixation probability for individual nodes in various update rules.

$$\mathbf{BD} - \mathbf{B} : F_{\{i\}} \leq r(r + \sum_j w_{ji})^{-1} \quad (10)$$

$$\mathbf{BD} - \mathbf{D} : F_{\{i\}} \leq \left(\sum_j \frac{w_{ji}}{r - rw_{ji} + w_{ji}} \right)^{-1} \quad (11)$$

$$\mathbf{DB} - \mathbf{B} : F_{\{i\}} \leq \sum_j rw_{ij}(1 - w_{ij} + rw_{ij})^{-1} \quad (12)$$

$$\mathbf{DB} - \mathbf{D} : F_{\{i\}} \leq r \sum_j w_{ij} \quad (13)$$

6. A Lower Bound for Mean Time to Fixation

Another important, although less-studied problem with respect to evolutionary graph theory is the mean time to fixation - the average time it takes for a mutant to take over the population. Closely related to this problem are mean time to extinction (average time for the resident to take over) and mean time to absorption (average time for either mutant or resident to take over). This has been previously studied under the original Moran process for well mixed populations [2] as well as some special cases of graphs [5]. However, to our knowledge, a general method to compute these quantities (without resorting to the use of simulation) have not been previously studied. Here we take a “first step” toward developing such a method by showing how the techniques introduced in this paper can be used to compute a non-trivial lower bound for mean time to fixation (and easily modified to bound mean time to extinction and absorption).

Let $F_{t|C}$ be the probability of fixation at time t . Therefore, $F_{t|C} - F_{t-1|C}$ is the probability of entering fixation at time t . The symbol t_C is the mean time. By the results of [2], we have the following:

Theorem 6.

$$t_C = \frac{1}{F_C} \sum_{t=1}^{\infty} t \cdot (F_{t|C} - F_{t-1|C})$$

Our key intuition is noticing that at each time step t , $F_{t|C} \leq \min_i P_{i,t}$. From this, we use the accounting method to provide a rigorous proof for the following theorem that provides a non-trivial lower-bound for the mean time to fixation. This result can be easily modified for mean time to extinction and absorption as well.

Theorem 7. $\frac{1}{F_C} \sum_{t=1}^{\infty} t \cdot (P_{\min,t} - P_{\min,t-1}) \leq \frac{1}{F_C} \sum_{t=1}^{\infty} t \cdot (F_{t|C} - F_{t-1|C})$ Where $P_{\min,t} = \min_i P_{i,t}$.

7. Algorithm and Experimental Evaluation

We leverage the finding of the previous sections in Algorithm 1. As described earlier, our method has found the exact fixation probability when all the probabilities in $\bigcup_i \{P_{i,t}\}$ (represented in the pseudo-code as the vector \mathbf{p}) are equal. We use Equation 4 to provide a convergence criteria based on value ϵ , which we can prove to be the tolerance for the fixation probability.

Proposition 2. *Algorithm 1 returns the fixation probability F_C within $\pm\epsilon$.*

Our novel method for computing fixation probabilities on strongly connected directed graphs allows us to compute near-exact fixation probabilities within a desired tolerance. The running time of the algorithm is highly dependent on how fast the vertex probabilities converge. In this section we experimentally evaluate how the vertex probabilities in our algorithms converge. We also provide results from comparison experiments to support the claim that *Algorithm 1-ACC* finds adequate fixation probabilities order of magnitudes faster than Monte Carlo simulations. We also show how the algorithm can be used to study the expected number of mutants as well as bound mean time to fixation.

7.1. Convergence of Vertex Probabilities

We ran our algorithm to compute fixation probabilities on randomly weighted and strongly connected directed graphs in order to experimentally evaluate the convergence of the vertex probabilities. We generated the graphs to be scale-free using the standard preferential attachment growth model [3] and randomly assigned an initial mutant node. We replaced all edges in the graph given by the growth model with two directed edges and then randomly assigned weights to all the edges.

Algorithm 1 - Our Novel Solution Method to Compute Fixation Probabilities

Input: Evolutionary Graph $\langle N, V, W \rangle$, configuration $C \subseteq V$, natural number $R > 0$, and real number $\epsilon \geq 0$.

Output: Estimate of fixation probability of mutant.

```
1:  $p_i$  is the  $i$ th position in vector  $p$  corresponding with vertex  $i \in V$ .
2: Set  $p_i = 1$  if  $i \in C$  and  $p_i = 0$  otherwise. {As per Proposition 1}
3:  $\mathbf{q} \leftarrow \mathbf{p}$  { $\mathbf{q}$  will be  $\mathbf{p}$  from the previous time step.}
4:  $\tau \leftarrow 1$ 
5: while  $\tau > \epsilon$  do
6:   for  $i \in V$  {This loop carries out the calculation as per Equation 3}
    do
8:    $sum \leftarrow 0$ 
9:    $\mathbf{m} \leftarrow \{j \in V | w_{ji} > 0\}$ 
10:  for  $j \in \mathbf{m}$  do
11:     $sum = sum + w_{ji} \cdot (\mathbf{q}_j - \mathbf{q}_i)$ 
12:  end for
13:   $\mathbf{p}_i \leftarrow \mathbf{q}_i + 1/N \cdot sum$ 
14: end for
15:  $\mathbf{q} \leftarrow \mathbf{p}$ 
16:  $\tau \leftarrow (1/2) \cdot (\max p - \min p)$  {Ensures error bound based on Equation 4}
17: end while
18: return  $(\min p) + \tau$ 
```

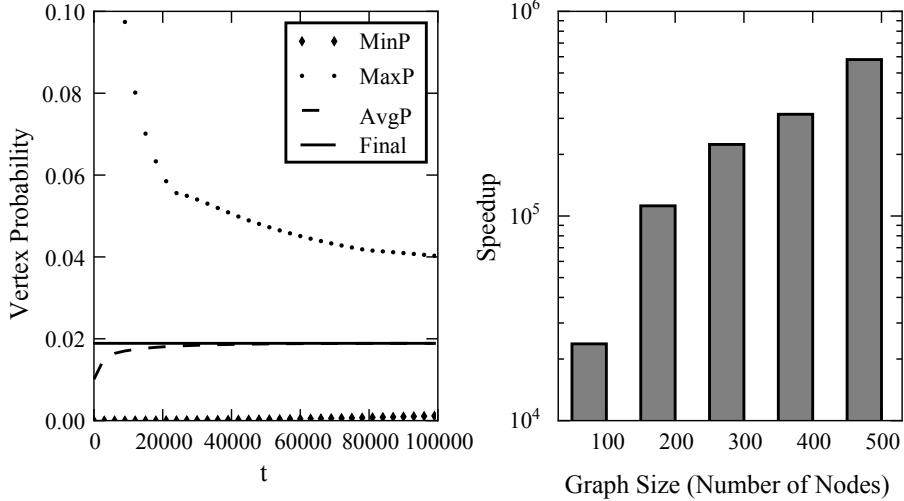


Figure 1: Left: Convergence of the minimum (MinP), maximum (MaxP), and average (AvgP) of vertex probabilities towards the final fixation probability as a function of our algorithm’s iterations t for a graph of 100 nodes. Right: Average speedup (on a log scale) for finding fixation probabilities achieved by our algorithm vs Monte Carlo simulation for graphs of different sizes.

To compare Algorithm 1 with the Monte Carlo approach, we should set the parameter R in that algorithm to be comparable with ϵ in Algorithm 1. As ϵ is the provable error of a solution to Algorithm 1. Based on the commonly-accepted definition of estimated standard error from statistics, we can obtain the estimated standard error for the solution returned by Monte Carlo approach with the following expression (where R is the number of simulation runs).

$$\sqrt{\frac{F_C(1 - F_C)}{R - 1}} \quad (14)$$

We can use Equation 14 to estimate the parameter R for the Monte Carlo approach as follows. We set ϵ equal to the estimated standard error as per Expression 14 and manipulate it algebraically. This gives us $R \approx \frac{S(S-1)}{\epsilon^2} + 1$ where S is the solution to Algorithm 1, ϵ is the input parameter for Algorithm 1 and R is the number of simulation runs in the Monte Carlo approach that we estimate to provide a comparable error bound. We also note, that as the vertex probabilities converge, the standard deviation of the p vector in Algorithm 1 could be a potentially faster convergence criteria.

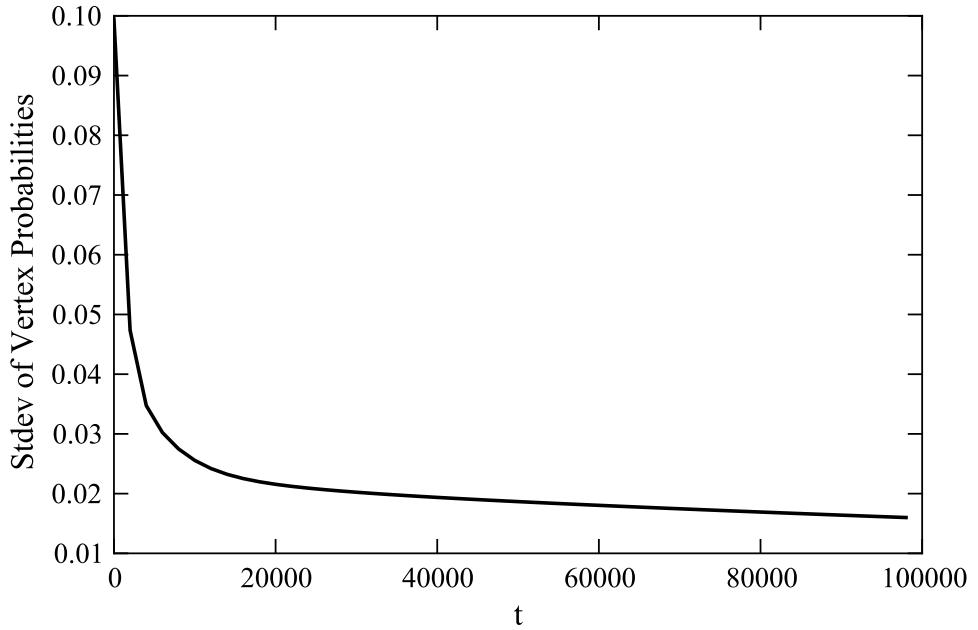


Figure 2: Standard deviation of vertex probabilities as a function of our algorithm’s iterations for the same 100 node graph of Figure 1 (left).

Note that using standard deviation of p and returning the average vertex probability would no longer provide us of the guarantee in Proposition 2, however it may provide good results in practice. The modifications to the algorithm would be as follows: line 15 would be $\tau \leftarrow \text{st.dev}(p)$ and line 17 would be **return** $\text{avg}(p)$. We will refer to this as *Algorithm 1 with alternate convergence criteria* or *Algorithm 1-ACC* for short.

Figure 1 (left) shows the convergence of the minimum, maximum, and the average of vertex probabilities towards the final fixation probability value for a small graph of 100 nodes. We can observe that the average converges to the final value at a logarithmic rate and much faster than the minimum and maximum vertex probability values. This suggests that while *Algorithm 1-ACC* does not give the same theoretical guarantees as *Algorithm 1*, it is much preferable for speed since the minimum and maximum vertex probabilities take much longer to converge to the final solution than the average. The fact that the average of the vertex probabilities is much preferable as a fast estimation of fixation probabilities is supported by the logarithmic decrease of the standard deviation of vertex probabilities (see Figure 2). Convergences

for other and larger graphs are not shown here but are qualitatively similar to the relative convergences shown in the provided graphs.

7.2. Speed Comparison to Monte Carlo Simulation

In order to compare our method’s speed compared to the standard Monte Carlo simulation method, we must determine how many iterations our algorithm must be run to find a fixation probability estimate comparable to that of the Monte Carlo approach. Thankfully, as we have seen, we can get a standard error on the fixation probability returned by the Monte Carlo approach as per Equation 14. While we did not theoretically prove anything about how smoothly fixation probabilities from our methods approach the final solution, the convergences of the average and standard deviation as shown above strongly suggest that estimates from our method approach the final solution quite gracefully. In fact, in the following experiments, once our method has arrived at a fixation probability estimate within the standard error of simulations, the estimate never again fell outside the window of standard error (although the estimate did not always approach the final estimate monotonically). This is in stark contrast to Monte Carlo simulations, from which estimations can vary greatly before the method has completed enough single runs to achieve a good probability estimate.

We generated a number of randomly weighted and strongly connected directed graphs of various sizes on which we compare our solution method to Monte Carlo approximation of fixation probabilities. The graphs were generated as in our convergence experiments. For each graph of a different size, we generated a number of different initial mutant configurations. We found fixation probabilities both using Monte Carlo estimation with 2000 simulation runs and our direct solution method, terminating when we have reached within the standard error of the Monte Carlo estimation. Since the average vertex probability proved to be such a good fast estimate of the true fixation probability, we used *Algorithm 1-ACC*.

Figure 1 (right) shows the speedup our solution provides over Monte Carlo simulation. Here speedup is defined as the ratio of the time it takes for simulations to complete over the time it takes our algorithm to find a fixation probability within the standard deviation. The often extremely low number of iterations needed by our algorithm to find fixation probabilities within the standard error of simulations may prompt the concern that the probabilities fall within this window so soon by mere chance. However, our experiments have shown that the fixation probability estimation found by

our algorithm at each iteration approaches the final fixation probability after termination smoothly at a logarithmic rate, asymptotically approaching the true fixation probability. While in this case the fixation probability estimate slightly crosses over the true fixation probability and then slowly approaches it again, none of the fixation probability estimates from our algorithm exited the window of standard error (from simulations) once they entered it.

We can observe from our experiments that computing fixation probabilities using Monte Carlo simulations showed to be a very time-expensive process, highlighting the need for faster solution methods as the one we have presented. Especially for larger graph sizes, the time complexity of our solution to achieve similar results to Monte Carlo simulation has shown to be orders of magnitude smaller than the standard method.

7.3. Monitoring the Expected Number of Mutants

As observed in Section 3, our method not only allows for the calculation of the fixation probability of a mutant, but also allows us to study how the expected number of mutants change over time. In this section, we present experimental results exploring the trajectory of the expected number of mutants over time on various undirected/unweighted graphs and under different initial mutant placement conditions.

First, we note that the expected number of mutants (as time approaches infinity) in an unweighted/undirected graph with respect to a single initially infected vertex i can be computed by modifying the result of [6] (for BD updating) to obtain the following.

$$\lim_{t \rightarrow \infty} \mathbf{Ex}_{\{i\}}^{(t)} = \frac{1}{k_i \langle k^{-1} \rangle} \quad (15)$$

Where $\langle k^{-1} \rangle$ is the average inverse of the degree for the graph. Hence, we can determine whether a node amplifies or suppresses selection by observing if $\lim_{t \rightarrow \infty} \mathbf{Ex}_{\{i\}}^{(t)}$ is greater or less than 1 respectively: if $k_i < \frac{1}{\langle k^{-1} \rangle}$ selection is amplified and if $k_i > \frac{1}{\langle k^{-1} \rangle}$ it is suppressed. We have used our algorithm to compare the trajectory of the expected number of mutants over time when the initial mutant is placed on amplifiers vs. suppressors under different graph topologies and BD updating. We note that similar comparisons can be obtained with our algorithm for the other update rules. We also note that by Theorem 5, an amplifier for BD (with no bias) will also be an amplifier for the (biased) BD-B and BD-D where $r > 1$.

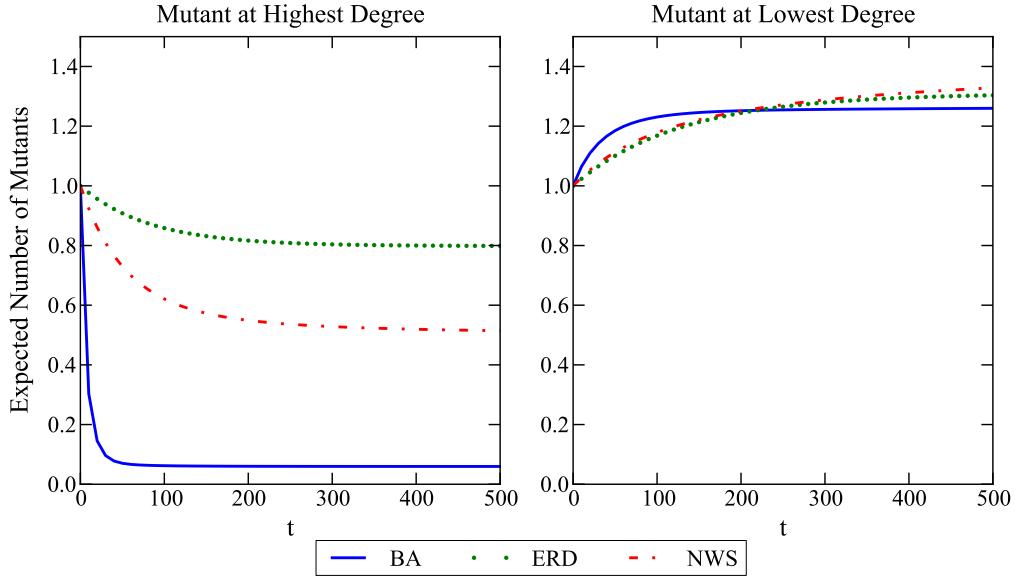


Figure 3: Expected number of mutants over time starting with a single mutant placed on a graph for Barabási-Albert preferential attachment (BAR), Erdős-Rényi (ERD), and Newmann-Watts-Strogatz small world (NWS) graphs. Lines are averages over 50 random graphs of each type. In the left graph, mutants are placed at the highest degree nodes, which are suppressors. In the right graph, mutants are placed at lowest degree nodes, which are amplifiers.

Figure 3 shows the trajectories of the expected number of mutants over time on random [3] preferential attachment (BAR), [10] (ER), and [19] small world graphs (NWS), each for when the initial mutant is placed on a suppressor (highest degree node of graph) and amplifier (lowest degree node of graph). Graphs are all of equal size at 100 nodes. We note that the highest degree nodes are especially strong suppressors on BAR graphs, less so for NWS graphs, and even less so for ER graphs. This makes sense when one considers the degree distribution of the different graph topologies, which are scale-free or power-law ($P(k) \sim k^{-3}$) for BR, roughly Poisson-shaped for NWS, and relatively uniform for ER graphs. For lowest degree amplifiers, the expected number of mutants grows faster early on in Barabási-Albert graphs, but it plateaus earlier than and is eventually surpassed by the slower growing expected number of mutants in the Erdős-Rényi, and Newmann-Watts-Strogatz graphs. Such insights into the evolutionary process may be

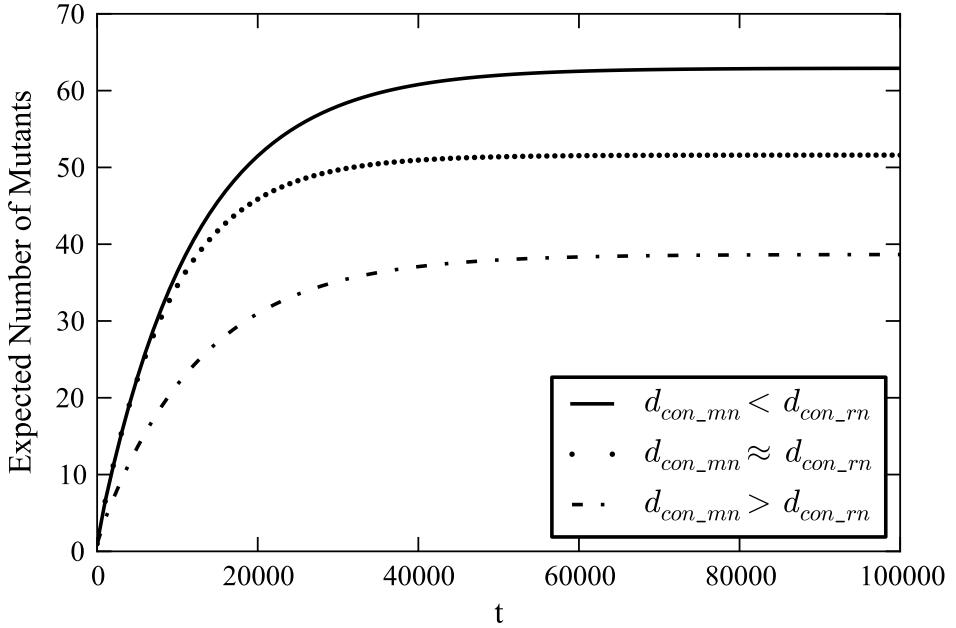


Figure 4: Expected number of mutants over time for an Erdős-Rényi graph of 100 nodes, with an extra mutant node (mn) and resident node(rn) with directed edges to con_mn and con_rn respectively. The value that the expected number of mutants converges to depends on the relative degrees of con_mn and con_rn , as shown in the legend.

crucial in applications, e.g. when one may be more interested in achieving highest number of mutants in a short amount of time rather than highest number of mutants as $t \rightarrow \infty$ or vice versa.

Finally, thus far we have only considered strongly connected graphs in which the vertex probabilities converge as $t \rightarrow \infty$, but this is not the case for some non-strongly connected graphs. We have thus also investigated the expected number of mutants over time for some simple cases of such graphs. Consider a random graph that is strongly connected, and then have a resident node (rn) and mutant node (mn) connected with only directed edges into the strongly connected graph. Clearly, the vertex probabilities cannot converge, since $\forall t, P_{mn,t} = 1$ and $P_{rn,t} = 0$. Our experimental results in Figure 4 show however that while the vertex probabilities do not converge, the value for the expected number of mutants given by our algorithm seems to converge. What value the expected number of mutants converges to depends on the relative degrees of the nodes that the mutant node mn and resident node

rn connect to. We shall call these nodes con_mn and con_rn , respectively. If $k_{con_mn} \approx k_{con_rn}$, the expected value of mutants converges at around 50% of the graph's nodes. If $k_{con_mn} > k_{con_rn}$, the expected value of mutants is less than 50% of the graph's nodes, and conversely, if $k_{con_mn} < k_{con_rn}$, it is greater. These results are intuitive because lower degree nodes are better spreaders under BD updating. These results are also interesting because the expected value converges - even though the graphs are not strongly connected. By an examination of Equation 6, this convergence is possible. However, we have not proven that convergence always occurs. An interesting direction for future work is to identify under what conditions will the expected number of mutants converges in a non-strongly connected graph.

7.4. Experimentally Computing the Lower Bound of the Mean Time to Fixation

We also performed experiments to examine the lower bound on mean time to fixation (discussed in Section 6) as compared to the average fixation time determined from simulation run. In doing so, we were able to confirm the lower-bound experimentally. We were able to use Algorithm 1-ACC to compute the lower bound with a few changes (noted in the supplement).

We generated random (ER) graphs of size 10, 20, 50 and 100 nodes, creating five different graphs for each number of nodes. The graphs were generated as in our convergence experiments, and our comparison to Monte Carlo testing are shown in Figure 5 where we demonstrate experimentally that our algorithm produces a lower bound. Our algorithm was run until the standard deviation of fixation probabilities for all vertices was 2.5×10^{-6} . The Monte Carlo simulations were each set at 10,000 runs.

8. Application: Competition Among Neural Axons

In recent work, [26] created a model for synaptic competition based on death-birth updating under neutral drift. They noted that the model aligns well with their empirical observations. In the model, the graph represents a synaptic junction and the nodes represent sites in the junction. For every two adjacent sites in the synaptic junction, there is an undirected edge between the corresponding two nodes in the graph. Hence, in- and out- degrees of each node are the same. Initially, there are K different axon types located in the junction configured in a manner where all sites are initially occupied by one axon type. At each time step, an axon occupying one of the sites is

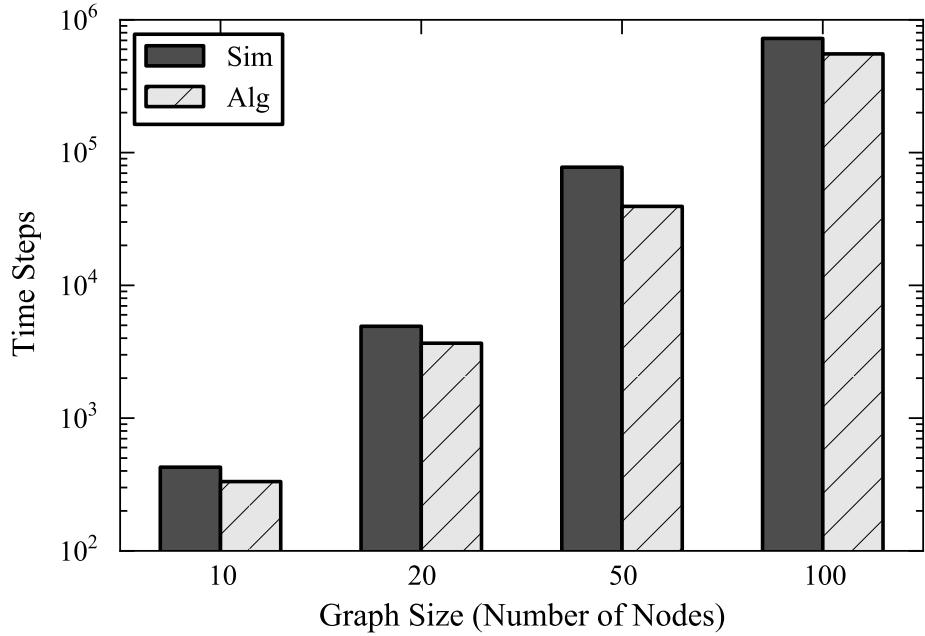


Figure 5: Mean-time-to-fixation comparison between algorithm and simulation. Note that the y-axis is a logarithmic scale.

eliminated - making the site open. The selection of the axon for elimination (death) is with a uniform probability. Hence, there is no bias in this model. Following the elimination of an axon, an adjacent axon grows into the site. The adjacent axon is selected with a uniform probability of the eliminated axon's neighbors. Hence, based on the results of this paper, we can provide the following insights into synaptic competition.

1. After t axons are eliminated,¹ the probability of any site being occupied by an axon of a certain type can be calculated directly by Theorem 8. Even though there are K axon types, this theorem still applies as it only considers the probability of a node being a mutant (resp. a site being a one of the K axon types).
2. Using point 1 above, we can determine the expected number of axons of a given type after t axons being eliminated.

¹Note that the number of axons eliminated corresponds directly to the number of timesteps in the model.

3. After t axons are eliminated, the probability of any set of sites being occupied by a certain axon type is simply the sum of the probabilities of the individual sites being occupied by that axon. As a result, the fixation probability is additive.
4. Leveraging point 3 above combined with an easy modification of the result of Broom et al. [6] for the BD model, the fixation probability of an axon originating at site i is $\frac{k_i}{2\Theta}$ where k_i is the number of sites adjacent to site i (hence the degree of node i in the corresponding graph) and Θ is the total number of adjacencies in the synapse (hence, half the number of directed edges in the corresponding graph).
5. Based on item 4 above and the results from Section 3, we can conclude that for a given axon type (let's call it "axon type A") occupying a set of sites, that if the average adjacencies of those sites is greater than (resp. less than) the overall average adjacencies for the sites in the entire synaptic junction, then as the number of eliminated axons approaches infinity, we can expect the number of axon type A in the synaptic junction will increase (resp. decrease) in expectation.
6. We can directly apply Theorem 7 to find a lower-bound on the number of eliminated axons before fixation occurs.

We note that the results stated above are either precise mathematical arguments or calculations that can be found exactly with a deterministic algorithm. They are not theoretical approximations and do not rely on simulation. As such is the case, we can make more precise statements about synaptic competition (given the model) and can avoid the variance that accompanies simulation results. Insights such as these may lead to future biological experiments.

9. Related Work

Evolutionary graph theory was originally introduced in [14]. Previously, we have compiled a comprehensive review [24] for a general overview of the work in this exciting new area.

While most work dealing with evolutionary graphs rely on Monte Carlo simulation, there are some good analytical approximations for the undirected/unweighted case based on the degree of the vertices in question. Antal et al. [1] use the mean-field approach to create these approximations for the undirected/unweighted case. Broom et al. [6] derive an exact analytical result for the undirected/unweighted case in neutral drift, which agrees

with the results of Antal et al. They also show that fixation probability is additive in that case (a result which we extend in this paper using a different proof technique). However, the results of [16] demonstrate that mean-field approximations break down in the case of weighted, directed graphs. [15] also studied weighted, directed graphs, but does so by using Monte Carlo simulation. [22] derive exact computation of fixation probability through means of linear programming. However, that approach requires an exponential number of both constraints and variables and is intractable. The recent work of [27] introduces a parameter called graph determinacy which measures the degree to which fixation or extinction is determined while starting from a randomly chosen initial configuration. This property is then used to analyze some special cases of evolutionary graphs under birth-death updating. There has been some work on algorithms for fixation probability calculation that rely on a randomized approach [4, 9]. [4] present a heuristic technique for speeding up Monte Carlo simulations by early termination while [9] present utilize simulation runs in a fully-polynomial randomized approximation scheme. However, our framework differs in that it does not rely on simulation at all and provides a deterministic result. Further, our non-randomized approach also allows for additional insights into the evolutionary process - such as monitoring the expected number of mutants as a function of time. Recently, [12] study the related problem of determining the probability of fixation given a single, randomly placed mutant in the graph where the vertices are “islands” and there are many individuals residing on each island in a well-mixed population. They use quasi-fixed points of ODE’s to obtain an approximation of the fixation probability and performed experiments with a maximum of 5 islands (vertices) containing 50 individuals each. This continuous approximation provides the best results when the number of individuals in each island is much larger than the number of islands. As the problem of this paper can be thought of as a special case where each island has just one individual, it seems unlikely that the approximation of Houchmandzadeh and Vallade’s approach will hold here.

Some of the results in this paper were previously presented in conferences by the authors [23, 17]. The analysis and experiments concerning the expected number of mutants at a given time, the extension of the framework for other update rules (beyond birth-death), the use of the framework for the case of $r > 1$, and the neurology applications are all new results appearing for the first time in this paper.

10. Conclusion

In this paper, we introduced a new approach to deal with problems relating to evolutionary graphs that rely on “vertex probabilities.” Our presented analytical method is the first deterministic method to compute fixation probability and provides a number of novel uses and results for EGT problems:

- Our method can be used to solve for the fixation probability under neutral drift orders of magnitude faster than Monte Carlo simulations, which is currently the presiding employed method in EGT studies. We have extended the method to all of the commonly used update processes in EGT. The special case of neutral drift is not only of interest in the literature [6, 15] but also it has been applied to problems in neurology [26].
- While the presented method is currently constrained to the case of neutral drift, we have demonstrated how it can inform cases of non-neutral drift by using it to provide both a lower and upper bound for this case. Combined with our analytical method’s speed, this means that it can be used to acquire useful knowledge to guide general EGT studies interested in the case of advantageous mutants.
- We have shown how our analytical method can be used to calculate a non-trivial lower bound to the mean time to fixation, providing a first step for a general method to computing this and related quantities that is lacking in the current literature.
- We have shown how our method can be used to calculate deterministically the expected number of mutants, which is useful for applications that require predictions on the number of mutants in the population under a specific finite time horizon. We have also provided results on the expected number of mutants on different common graph topologies, showing differences in the growth trajectories of amplifiers and suppressors on these different topologies. These results may prove highly significant in the recent application of EGT to distributed systems [13] where the problem of information diffusion is considered among computer systems. In such a domain, it may be insufficient to guarantee fixation in the limit of time - which may be impractical - but rather to make guarantees on the outcome of the process after a finite amount of time.

- Finally, we have shown how our method can provide insight when applied to the problem of synaptic competition in neuroscience.

Though evolutionary graph theory is still a relatively new research area, it is actively being studied in a variety of disciplines [14, 24, 28, 25, 20, 26, 13]. We believe that more real-world applications will appear as this area gains more popularity. As illustrated by recent work [26, 13], experimental scientists with knowledge of EGT may be more likely to recognize situations where the model may be appropriate. As these cases arise, deterministic methods for addressing issues related to EGT may prove to be highly useful. However, this paper is only a starting point - there are still many important directions for future work. Foremost among such topics are scenarios where the topology of the graph also changes over time or where additional attributes of the nodes/edges in the graph affect the dynamics.

Acknowledgments

P.S. is supported by ARO projects 611102B74F and 2GDATXR042 as well as OSD project F1AF262025G001. P.R. is supported by ONR grant W911NF0810144. P.S. would like to thank Stephen Turney (Harvard University) for several discussions concerning his work. The opinions in this paper are those of the authors and do not necessarily reflect the opinions of the funders, the U.S. Military Academy, the U.S. Army, or the U.S. Navy.

References

- 1 Antal, T., Redner, S., Sood, V., 2006. Evolutionary dynamics on degree-heterogeneous graphs. *Physical Review Letters* 96 (18), 188104.
- 2 Antal, T., Scheuring, I., 2006. Fixation of strategies for an evolutionary game in finite populations. *Bulletin of Mathematical Biology* 68, 1923–1944.
- 3 Barabási, A., Albert, R., 1999. Emergence of scaling in random networks. *science* 286 (5439), 509–512.
- 4 Barbosa, V. C., Donangelo, R., Souza, S. R., Oct 2010. Early appraisal of the fixation probability in directed networks.

5 Broom, M., Hadjichrysanthou, C., Rychtář, J., 2009. Evolutionary games on graphs and the speed of the evolutionary process. *Proceedings of the Royal Society A*.

6 Broom, M., Hadjichrysanthou, C., Rychtář, J., Stadler, B. T., Apr. 2010. Two results on evolutionary processes on general non-directed graphs. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* 466 (2121), 2795–2798.

7 Broom, M., Rychtář, J., May 2009. An analysis of the fixation probability of a mutant on special classes of non-directed graphs. *Proceedings of the Royal Society A* 464, 2609–2627.

8 Broom, M., Rychtář, J., Stadler, B., 2011. Evolutionary dynamics on graphs - the effect of graph structure and initial placement on mutant spread. *Journal of Statistical Theory and Practice* 5 (3), 369–381.

9 Díaz, J., Goldberg, L., Mertzios, G., Richerby, D., Serna, M., Spirakis, P., Jan. 2012. Approximating Fixation Probabilities in the Generalized Moran Process. In: *Proceedings of the ACM-SIAM Symposium on Discrete Algorithms (SODA)*, Kyoto, Japan. ACM.

10 Erdős, P., Rényi, A., 1960. On the evolution of random graphs. Akad. Kiadó.

Hagberg et al. Hagberg, Aric A. and Schult, Daniel A. and Swart, Pieter J.. Exploring network structure, dynamics, and function using NetworkX. In *Proceedings of the 7th Python in Science Conference (SciPy2008)*, pages 11–15, Pasadena, CA USA, August 2008.

12 Houchmandzadeh, B., Vallade, M., July 2011. The fixation probability of a beneficial mutation in a geographically structured population. *New Journal of Physics* 13 (7), 073020.
URL <http://stacks.iop.org/1367-2630/13/i=7/a=073020>

13 Jiang, C., Chen, Y. C., Liu, K. J. R., Dec 2012. Distributed Adaptive Networks: A Graphical Evolutionary Game-Theoretic View. arXiv: 1212.1245.

14 Lieberman, E., Hauert, C., Nowak, M. A., 2005. Evolutionary dynamics on graphs. *Nature* 433 (7023), 312–316.

15 Masuda, N., 2009. Directionality of contact networks suppresses selection pressure in evolutionary dynamics. *Journal of Theoretical Biology* 258 (2), 323 – 334.

16 Masuda, N., Ohtsuki, H., 2009. Evolutionary dynamics and fixation probabilities in directed networks. *New Journal of Physics* 11 (3), 033012 (15pp).

17 Moores, G., Shakarian, P., 2012. A fast and deterministic method for mean time to fixation in evolutionary graphs. Presented at INSNA Sunbelt XXXII, Redondo Beach, CA.

18 Moran, P., 1958. Random processes in genetics. *Mathematical Proceedings of the Cambridge Philosophical Society* 54 (01), 60–71.

19 Newman, M., Watts, D., 1999. Renormalization group analysis of the small-world network model. *Physics Letters A* 263 (4-6), 341–346.

20 Pacheco, J. M., Traulsen, A., Nowak, M. A., December 2006. Active linking in evolutionary games. *Journal of Theoretical Biology* 243 (3), 437–443.

21 Paley, C. J., Taraskin, S. N., Elliott, S. R., 2007. Temporal and dimensional effects in evolutionary graph theory. *Physical Review Letters* 98, 098103.

22 Rychtář, J., Stadler, B., Winter 2008. Evolutionary dynamics on small-world networks. *International Journal of Computational and Mathematical Sciences* 2 (1).

23 Shakarian, P., Roos, P., Nov. 2011. Fast and deterministic computation of fixation probability in evolutionary graphs. In: CIB '11: The Sixth IASTED Conference on Computational Intelligence and Bioinformatics. IASTED.

24 Shakarian, P., Roos, P., Johnson, A., 2012. A review of evolutionary graph theory with applications to game theory. *Biosystems* 107 (2), 66 – 80.

URL <http://www.sciencedirect.com/science/article/pii/S0303264711001675>

- 25 Sood, V., Antal, T., Redner, S., 2008. Voter models on heterogeneous networks. *Physical Review E* 77 (4), 041121.
- 26 Turney, S., Lichtman, J., June 2012. Reverseing the outcome of synapse elimination at developing neuromuscular junction *in vivo*: Evidence for synaptic competition and its mechanism. *PLoS Biology* 10.
- 27 Voorhees, B., 2012. Birth-Death Fixation Probabilities for Structured Populations. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences* (accepted).
- 28 Zhang, P. A., Nie, P. Y., Hu, D. Q., Zou, F. Y., 2007. The analysis of bi-level evolutionary graphs. *Biosystems* 90 (3), 897–902.

Supplementary Material

11. Notes

Throughout this supplement, we will use an extended notation. Fixation probability given initial configuration C is denoted F_C . For vertex i at time t , we denote this as $\Pr(\mathbf{M}_i^{(t)})$. We will use $\mathbf{S}_i^{(t)}$ to denote the event that vertex i was selected for reproduction and $\mathbf{R}_{ij}^{(t)}$ to denote the event of i replacing j . We will often use conditional probabilities. For example, $\Pr(\mathbf{M}_i^{(t)}|C^{(0)})$ is the probability that v_i is a mutant given the initial set C of mutants. Throughout this supplement, unless noted otherwise, all of our probabilities will be conditioned on $C^{(0)}$. We will drop it for ease of notation with the understanding that some set C of V were mutants at $t = 0$. Hence, $\Pr(\mathbf{M}_i^{(t)}) = \Pr(\mathbf{M}_i^{(t)}|C^{(0)})$.

12. Proof of Theorem 1

$$\forall i, \quad \lim_{t \rightarrow \infty} \Pr(\mathbf{M}_i^{(t)}|C^{(0)}) = F_C$$

Proof. Consider the following definition property of $\Pr(\mathbf{M}_i^{(t)}|C^{(0)})$

$$\Pr(\mathbf{M}_i^{(t)}|C^{(0)}) = \sum_{\substack{V' \in 2^V \\ s.t. \ v_i \in V'}} \Pr(V'^{(t)}|C^{(0)}) \quad (16)$$

We note that as time approaches infinity, for all $V' \in 2^V - \emptyset - V$ we have $\Pr(V'^{(t)}|C^{(0)}) = 0$. As $v_i \notin \emptyset$, the statement follows. *Q.E.D.*

13. Proof of Theorem 2

$$\Pr(\mathbf{M}_i^{(t)}) =$$

$$\Pr(\mathbf{M}_i^{(t-1)}) + \sum_{(v_j, v_i) \in E} w_{ji} \cdot \Pr(\mathbf{M}_j^{(t-1)}) \cdot \Pr(\mathbf{S}_j^{(t)}|\mathbf{M}_j^{(t-1)}) - w_{ji} \cdot \Pr(\mathbf{M}_i^{(t-1)}) \cdot \Pr(\mathbf{S}_j^{(t)}|\mathbf{M}_i^{(t-1)})$$

Where $\mathbf{S}_i^{(t)}$ is true iff v_i is selected for reproduction at time t .

Proof. Note we use the variable $\mathbf{R}_j^{(t)} i$ is true iff v_j replaces v_i at time t .

CLAIM 1:

$$\begin{aligned} \Pr(\mathbf{M}_i^{(t)}) &= \Pr(\mathbf{M}_i^{(t-1)} \wedge \bigwedge_{(v_j, v_i) \in E} \neg \mathbf{S}_j^{(t)}) + \sum_{(v_j, v_i) \in E} \Pr(\mathbf{S}_j^{(t)} \wedge \mathbf{R}_{ji}^{(t)} \wedge \mathbf{M}_j^{(t-1)}) + \\ &\quad \sum_{(v_j, v_i) \in E} \Pr(\mathbf{S}_j^{(t)} \wedge \neg \mathbf{R}_{ji}^{(t)} \wedge \mathbf{M}_i^{(t-1)}) \end{aligned}$$

This is shown by a simple examination of exhaustive and mutually exclusive events based on the original model of [14].

CLAIM 2:

$$\Pr(\mathbf{M}_i^{(t-1)} \wedge \bigwedge_{(v_j, v_i) \in E} \neg \mathbf{S}_j^{(t)}) = \Pr(\mathbf{M}_i^{(t-1)}) \cdot \left(1 - \sum_{(v_j, v_i) \in E} \Pr(\mathbf{S}_j^{(t)} | \mathbf{M}_i^{(t-1)}) \right)$$

(Proof of claim 2) By exhaustive and mutual exclusive events, we have the following.

$$\Pr(\mathbf{M}_i^{(t-1)} \wedge \bigwedge_{(v_j, v_i) \in E} \neg \mathbf{S}_j^{(t)}) = \Pr(\mathbf{M}_i^{(t-1)}) - \sum_{(v_j, v_i) \in E} \Pr(\mathbf{S}_j^{(t)} \wedge \mathbf{M}_i^{(t-1)})$$

By the definition of conditional probability, we have the following

$$\begin{aligned} \Pr(\mathbf{M}_i^{(t-1)} \wedge \bigwedge_{(v_j, v_i) \in E} \neg \mathbf{S}_j^{(t)}) &= \Pr(\mathbf{M}_i^{(t-1)}) - \sum_{(v_j, v_i) \in E} \left(\Pr(\mathbf{S}_j^{(t)} | \mathbf{M}_i^{(t-1)}) \cdot \Pr(\mathbf{M}_i^{(t-1)}) \right) \\ &= \Pr(\mathbf{M}_i^{(t-1)}) - \Pr(\mathbf{M}_i^{(t-1)}) \cdot \sum_{(v_j, v_i) \in E} \Pr(\mathbf{S}_j^{(t)} | \mathbf{M}_i^{(t-1)}) \\ &= \Pr(\mathbf{M}_i^{(t-1)}) \left(1 - \sum_{(v_j, v_i) \in E} \Pr(\mathbf{S}_j^{(t)} | \mathbf{M}_i^{(t-1)}) \right) \end{aligned}$$

The claim immediately follows.

CLAIM 3: For all edges (v_j, v_i) , we have the following.

$$\Pr(\mathbf{S}_j^{(t)} \wedge \mathbf{R}_{ji}^{(t)} \wedge \mathbf{M}_j^{(t-1)}) = w_{ji} \cdot \Pr(\mathbf{M}_j^{(t-1)}) \cdot \Pr(\mathbf{S}_j^{(t)} | \mathbf{M}_j^{(t-1)})$$

(Proof of claim 3) The following is a direct application of the definition of conditional probability.

$$\Pr(\mathbf{S}_j^{(t)} \wedge \mathbf{R}_{ji}^{(t)} \wedge \mathbf{M}_j^{(t-1)}) = \Pr(\mathbf{R}_{ji}^{(t)} \wedge \mathbf{M}_j^{(t-1)} | \mathbf{S}_j^{(t)}) \cdot \Pr(\mathbf{S}_j^{(t)})$$

From our model, we note that given even $\mathbf{S}_j^{(t)}$, the fitness of the nodes is not considered in determining if the event associated with $\mathbf{R}_{ji}^{(t)}$ is to occur. Hence, it follows that $\mathbf{M}_j^{(t-1)}$ is independent of $\mathbf{R}_{ji}^{(t)}$ given $\mathbf{S}_j^{(t)}$. As $\Pr(\mathbf{R}_{ji}^{(t)} | \mathbf{S}_j^{(t)}) = w_{ji}$, we have the following.

$$\begin{aligned} \Pr(\mathbf{S}_j^{(t)} \wedge \mathbf{R}_{ji}^{(t)} \wedge \mathbf{M}_j^{(t-1)}) &= \Pr(\mathbf{R}_{ji}^{(t)} | \mathbf{S}_j^{(t)}) \cdot \Pr(\mathbf{M}_j^{(t-1)} | \mathbf{S}_j^{(t)}) \cdot \Pr(\mathbf{S}_j^{(t)}) \\ &= w_{ji} \cdot \Pr(\mathbf{M}_j^{(t-1)} | \mathbf{S}_j^{(t)}) \cdot \Pr(\mathbf{S}_j^{(t)}) \end{aligned}$$

By Bayes Theorem, and that the model causes $\forall i \Pr(\mathbf{S}_i^{(t)}) > 0$, we have the following.

$$\begin{aligned} \Pr(\mathbf{S}_j^{(t)} \wedge \mathbf{R}_{ji}^{(t)} \wedge \mathbf{M}_j^{(t-1)}) &= w_{ji} \cdot \Pr(\mathbf{S}_j^{(t)} | \mathbf{M}_j^{(t-1)}) \cdot \frac{\Pr(\mathbf{M}_j^{(t-1)})}{\Pr(\mathbf{S}_j^{(t)})} \cdot \Pr(\mathbf{S}_j^{(t)}) \\ &= w_{ji} \cdot \Pr(\mathbf{M}_j^{(t-1)}) \cdot \Pr(\mathbf{S}_j^{(t)} | \mathbf{M}_j^{(t-1)}) \end{aligned}$$

The claim follows immediately.

CLAIM 4: For all edges (v_j, v_i) , we have the following.

$$\Pr(\mathbf{S}_j^{(t)} \wedge \neg \mathbf{R}_{ji}^{(t)} \wedge \mathbf{M}_i^{(t-1)}) = (1 - w_{ji}) \cdot \Pr(\mathbf{M}_i^{(t-1)}) \cdot \Pr(\mathbf{S}_j^{(t)} | \mathbf{M}_i^{(t-1)})$$

(Proof of claim 4) This mirrors claim 3.

(Proof of theorem) From claims 1-4, we have the following.

$$\Pr(\mathbf{M}_i^{(t)}) = \Pr(\mathbf{M}_i^{(t-1)}) \cdot \left(1 - \sum_{(v_j, v_i) \in E} \Pr(\mathbf{S}_j^{(t)} | \mathbf{M}_i^{(t-1)}) \right) + \quad (17)$$

$$\sum_{(v_j, v_i) \in E} \left(w_{ji} \cdot \Pr(\mathbf{M}_j^{(t-1)}) \cdot \Pr(\mathbf{S}_j^{(t)} | \mathbf{M}_j^{(t-1)}) \right) + \quad (18)$$

$$\sum_{(v_j, v_i) \in E} \left((1 - w_{ji}) \cdot \Pr(\mathbf{M}_i^{(t-1)}) \cdot \Pr(\mathbf{S}_j^{(t)} | \mathbf{M}_i^{(t-1)}) \right) \quad (19)$$

Which, after re-arranging some terms, gives us the statement of the theorem.
Q.E.D.

14. Proof of Theorem 3

When $r = 1$, if for some time t , $\forall i$, the value $\mathbf{Pr}(\mathbf{M}_i^{(t)})$ is the same, then $\mathbf{Pr}(\mathbf{M}_i^{(t)}) = F_C$.

Proof Sketch. Consider $\mathbf{Pr}(\mathbf{M}_i^{(t)}) = \mathbf{Pr}(\mathbf{M}_i^{(t-1)}) + \frac{1}{N} \sum_{(v_j, v_i) \in E} w_{ji} \cdot (\mathbf{Pr}(\mathbf{M}_j^{(t-1)}) - \mathbf{Pr}(\mathbf{M}_i^{(t-1)}))$ when for $t-1$, $\forall i, j$ we have $\mathbf{Pr}(\mathbf{M}_j^{(t-1)}) = \mathbf{Pr}(\mathbf{M}_i^{(t-1)})$. Clearly, in this case, the value for $\mathbf{Pr}(\mathbf{M}_i^{(t)}) = \mathbf{Pr}(\mathbf{M}_i^{(t-1)})$. As the probabilities of all vertices was the same at $t-1$, they remain so at t . Therefore, in this case, $\lim_{t \rightarrow \infty} \mathbf{Pr}(\mathbf{M}_i^{(t)}) = \mathbf{Pr}(\mathbf{M}_i^{(t-1)})$. QED

15. Proof of Inequality 4

For any time t , under neutral drift ($r = 1$),

$$\min_i \mathbf{Pr}(\mathbf{M}_i^{(t)}) \leq F_C \leq \max_i \mathbf{Pr}(\mathbf{M}_i^{(t)})$$

Proof. PART 1: For any time t , under neutral drift ($r = 1$), $F_C \leq \max_i \mathbf{Pr}(\mathbf{M}_i^{(t)})$.

We show that for each time step t , $\max_i \mathbf{Pr}(\mathbf{M}_i^{(t-1)}) \geq \max_i \mathbf{Pr}(\mathbf{M}_i^{(t)})$. Hence, by showing that, for any time t' , we have $\max_i \mathbf{Pr}(\mathbf{M}_i^{(t')}) \geq \lim_{t \rightarrow \infty} \max_i \mathbf{Pr}(\mathbf{M}_i^{(t)})$ which by allows us to apply Theorem 1 and obtain the statement of this theorem. Suppose BWOC that at time t we have $\max_\ell \mathbf{Pr}(\mathbf{M}_\ell^{(t-1)}) < \max_i \mathbf{Pr}(\mathbf{M}_i^{(t)})$.

Then we have:

$$\begin{aligned}
\max_{\ell} \mathbf{Pr}(\mathbf{M}_{\ell}^{(t-1)}) &< \frac{1}{N} \sum_{(v_j, v_i) \in E} w_{ji} \cdot \left(\mathbf{Pr}(\mathbf{M}_j^{(t-1)}) - \mathbf{Pr}(\mathbf{M}_i^{(t-1)}) \right) \\
&\quad + \mathbf{Pr}(\mathbf{M}_i^{(t-1)}) \\
&\leq \frac{1}{N} \sum_{(v_j, v_i) \in E} w_{ji} \cdot \left(\max_{\ell} \mathbf{Pr}(\mathbf{M}_{\ell}^{(t-1)}) - \mathbf{Pr}(\mathbf{M}_i^{(t-1)}) \right) \\
&\quad + \mathbf{Pr}(\mathbf{M}_i^{(t-1)}) \\
&= \frac{\sum_{(v_j, v_i) \in E} w_{ji}}{N} \left(\max_{\ell} \mathbf{Pr}(\mathbf{M}_{\ell}^{(t-1)}) - \mathbf{Pr}(\mathbf{M}_i^{(t-1)}) \right) \\
&\quad + \mathbf{Pr}(\mathbf{M}_i^{(t-1)}) \\
\max_{\ell} \mathbf{Pr}(\mathbf{M}_{\ell}^{(t-1)}) (1 - \frac{\sum_{(v_j, v_i) \in E} w_{ji}}{N}) &< \mathbf{Pr}(\mathbf{M}_i^{(t-1)}) (1 - \frac{\sum_{(v_j, v_i) \in E} w_{ji}}{N}) \\
\max_{\ell} \mathbf{Pr}(\mathbf{M}_{\ell}^{(t-1)}) &< \mathbf{Pr}(\mathbf{M}_i^{(t-1)})
\end{aligned}$$

Which is clearly a contradiction and completes this part of the proof.

PART 2: For any time t , under neutral drift ($r = 1$), $F_C \geq \min_i \mathbf{Pr}(\mathbf{M}_i^{(t)})$. We show that for each time step t , $\min_i \mathbf{Pr}(\mathbf{M}_i^{(t-1)}) \leq \min_i \mathbf{Pr}(\mathbf{M}_i^{(t)})$. Hence, by showing that, for any time t' , we have $\min_i \mathbf{Pr}(\mathbf{M}_i^{(t')}) \leq \lim_{t \rightarrow \infty} \max_i \mathbf{Pr}(\mathbf{M}_i^{(t)})$ which by allows us to apply Theorem 1 and obtain the statement of this theorem. Suppose BWOC that at time t we have $\min_{\ell} \mathbf{Pr}(\mathbf{M}_{\ell}^{(t-1)}) > \min_i \mathbf{Pr}(\mathbf{M}_i^{(t)})$.

Then we have:

$$\begin{aligned}
\min_{\ell} \Pr(\mathbf{M}_{\ell}^{(t-1)}) &> \frac{1}{N} \sum_{(v_j, v_i) \in E} w_{ji} \cdot \left(\Pr(\mathbf{M}_j^{(t-1)}) - \Pr(\mathbf{M}_i^{(t-1)}) \right) \\
&\quad + \Pr(\mathbf{M}_i^{(t-1)}) \\
&\geq \frac{1}{N} \sum_{(v_j, v_i) \in E} w_{ji} \cdot \left(\min_{\ell} \Pr(\mathbf{M}_{\ell}^{(t-1)}) - \Pr(\mathbf{M}_i^{(t-1)}) \right) \\
&\quad + \Pr(\mathbf{M}_i^{(t-1)}) \\
&= \frac{\sum_{(v_j, v_i) \in E} w_{ji}}{N} \left(\min_{\ell} \Pr(\mathbf{M}_{\ell}^{(t-1)}) - \Pr(\mathbf{M}_i^{(t-1)}) \right) \\
&\quad + \Pr(\mathbf{M}_i^{(t-1)}) \\
\min_{\ell} \Pr(\mathbf{M}_{\ell}^{(t-1)}) \left(1 - \frac{\sum_{(v_j, v_i) \in E} w_{ji}}{N} \right) &> \Pr(\mathbf{M}_i^{(t-1)}) \left(1 - \frac{\sum_{(v_j, v_i) \in E} w_{ji}}{N} \right) \\
\min_{\ell} \Pr(\mathbf{M}_{\ell}^{(t-1)}) &> \Pr(\mathbf{M}_i^{(t-1)})
\end{aligned}$$

Which is clearly a contradiction and completes this part of the proof. *Q.E.D.*

16. Proof of Theorem Theorem 4

When $r = 1$ for disjoint sets $C, D \subseteq V$, $F_C + F_D = F_{C \cup D}$.

Proof. Consider some time t and vertex v_i . Clearly, by Corollary 1, $\Pr(\mathbf{M}_i^{(t)})$ can be expressed as a linear combination of the form $\sum_{v_j \in V} (C_j \cdot \Pr(\mathbf{M}_j^{(0)}))$ where C_j is a coefficient. We note that these coefficients are the same regardless of the initial configuration of mutants that $\mathbf{M}_i^{(t)}$ is conditioned on. Hence, $\Pr(\mathbf{M}_i^{(t)} | C^{(0)})$ is this positive function with $\Pr(\mathbf{M}_j^{(0)}) = 1$ if $v_j \in C$ and 0 otherwise (see Proposition 3). Hence, for disjoint C, D , for any $v_i \in V$, we have $\Pr(\mathbf{M}_i^{(t)} | C^{(0)}) + \Pr(\mathbf{M}_i^{(t)} | D^{(0)}) = \Pr(\mathbf{M}_i^{(t)} | (C \cup D)^{(0)})$. The statement follows. *Q.E.D.*

17. Proof of Equation 9

$$\Pr(\mathbf{M}_i^{(t)}) = \left(1 - \frac{1}{N} \right) \cdot \Pr(\mathbf{M}_i^{(t-1)}) + \frac{1}{N \cdot k_{in}^{(i)}} \sum_{(v_j, v_i) \in E} \Pr(\mathbf{M}_j^{(t-1)})$$

Under death-birth dynamics with neutral drift ($r = 1$).

Proof. $\mathbf{D}_i^{(t)}$ and $\mathbf{B}_i^{(t)}$ are random variables associated with birth and death events for vertex v_i .

CLAIM 1: $\Pr(\mathbf{M}_i^{(t)}) = \Pr(\mathbf{M}_i^{(t-1)} \wedge \neg \mathbf{D}_i^{(t)}) + \sum_{(v_j, v_i) \in E} \Pr(\mathbf{D}_i^{(t)} \wedge \mathbf{B}_j^{(t)} \wedge \mathbf{M}_j^{(t-1)})$

Follows directly from exhaustive and mutually exclusive events.

CLAIM 2: $\Pr(\mathbf{M}_i^{(t-1)} \wedge \neg \mathbf{D}_i^{(t)}) = (1 - \frac{1}{N}) \cdot \Pr(\mathbf{M}_i^{(t-1)})$

By the definition of conditional probabilities, we have $\Pr(\neg \mathbf{D}_i^{(t)} | \mathbf{M}_i^{(t-1)}) \cdot \Pr(\mathbf{M}_i^{(t-1)})$. Also, we know the probability of a given node dying is always $1/N$. Hence, $\Pr(\neg \mathbf{D}_i^{(t)} | \mathbf{M}_i^{(t-1)}) = \Pr(\neg \mathbf{D}_i^{(t)}) = 1 - \frac{1}{N}$ and the claim follows.

CLAIM 3: For any $(v_j, v_i) \in E$, we have

$\Pr(\mathbf{D}_i^{(t)} \wedge \mathbf{B}_j^{(t)} \wedge \mathbf{M}_j^{(t-1)}) = \frac{1}{N \cdot k_{in}^{(i)}} \cdot \Pr(\mathbf{M}_j^{(t-1)})$

As both birth and death events occur independent of any node being a mutant at the previous time step, the definition of conditional probabilities gives us the following:

$$\Pr(\mathbf{D}_i^{(t)} \wedge \mathbf{B}_j^{(t)} \wedge \mathbf{M}_j^{(t-1)}) = \Pr(\mathbf{D}_i^{(t)} \wedge \mathbf{B}_j^{(t)}) \cdot \Pr(\mathbf{M}_j^{(t-1)}) \quad (20)$$

$$= \Pr(\mathbf{B}_j^{(t)} | \mathbf{D}_i^{(t)}) \cdot \Pr(\mathbf{D}_i^{(t)}) \cdot \Pr(\mathbf{M}_j^{(t-1)}) \quad (21)$$

From the model, we have the following:

$$\Pr(\mathbf{B}_j^{(t)} | \mathbf{D}_i^{(t)}) = 1/k_{in}^{(i)} \quad (22)$$

$$\Pr(\mathbf{D}_i^{(t)}) = 1/N \quad (23)$$

Hence, the claim follows. *QED Q.E.D.*

18. Proof of Equation 10

$$\Pr(\mathbf{M}_i^{(t)}) = \left(1 - \frac{k_{in}^{(i)}}{|E|}\right) \cdot \Pr(\mathbf{M}_i^{(t-1)}) + \frac{1}{|E|} \sum_{(v_j, v_i) \in E} \Pr(\mathbf{M}_j^{(t-1)})$$

Under link dynamics with neutral drift ($r = 1$).

Proof. Here $\mathbf{S}_{ij}^{(t)}$ is the random variable associated with the selection of edge (v_i, v_j) .

CLAIM 1: $\Pr(\mathbf{M}_i^{(t)}) = \Pr(\mathbf{M}_i^{(t-1)} \wedge \bigwedge_{(v_j, v_i) \in E} \neg \mathbf{S}_{ji}^{(t)}) + \sum_{(v_j, v_i) \in E} \Pr(\mathbf{S}_{ji}^{(t)} \wedge \mathbf{M}_j^{(t-1)})$

Follows directly from exhaustive and mutually exclusive events.

CLAIM 2: $\Pr(\bigwedge_{(v_j, v_i) \in E} \neg \mathbf{S}_{ji}^{(t)}) = 1 - \frac{k_{in}^{(i)}}{|E|}$

Clearly, we have $\Pr(\bigwedge_{(v_j, v_i) \in E} \neg \mathbf{S}_{ji}^{(t)}) = \Pr(\bigvee_{\{(v_\beta, v_\alpha) \in E | \beta \neq i\}} \mathbf{S}_{\beta\alpha}^{(t)})$. As there are $k_{in}^{(i)}$ incoming edges to v_i , we know that $\Pr(\bigvee_{\{(v_\beta, v_\alpha) \in E | \beta \neq i\}} \mathbf{S}_{\beta\alpha}^{(t)}) = 1 - \frac{k_{in}^{(i)}}{|E|}$, giving us the claim.

CLAIM 3: $\Pr(\mathbf{M}_i^{(t-1)} \wedge \bigwedge_{(v_j, v_i) \in E} \neg \mathbf{S}_{ji}^{(t)}) = \left(1 - \frac{k_{in}^{(i)}}{|E|}\right) \cdot \Pr(\mathbf{M}_i^{(t-1)})$

For any α, β , the random variable $\mathbf{S}_{\alpha\beta}^{(t)}$ is independent from $\mathbf{M}_i^{(t-1)}$. Hence, the claim immediately follows from this fact and claim 2.

CLAIM 4: $\Pr(\mathbf{S}_{ji}^{(t)} \wedge \mathbf{M}_j^{(t-1)}) = \frac{1}{|E|} \cdot \Pr(\mathbf{M}_j^{(t-1)})$

As, by the definition of the model, $\Pr(\mathbf{S}_{ji}^{(t)} | \mathbf{M}_j^{(t-1)}) = \Pr(\mathbf{S}_{ji}^{(t)}) = \frac{1}{|E|}$, the claim follows directly from the definition of conditional probabilities. *QED*
Q.E.D.

19. Proof of Theorem 5

For a given set C , let $F^{(1)}(C)$ be the fixation probability under neutral drift and $F^{(r)}(C)$ be the fixation probability calculated using a mutant fitness $r > 1$. Then, under BD-B, BD-D, DB-B, DB-D, or LD dynamics, $F^{(1)}(C) \leq F^{(r)}(C)$.

Proof. First, some notation.

- We define an **interpretation**, $\mathbf{I} : 2^V \rightarrow [0, 1]$ as probability distribution over mutant configurations. Hence, for some \mathbf{I} we have $\sum_{V' \in 2^V} \mathbf{I}(V') = 1$.
- Next, we define a transition function that maps configurations of mutants to probabilities, $\chi : 2^V \rightarrow [0, 1]$ where for any $C \in 2^V$, $\sum_{C' \in 2^V} \chi(C, C') = 1$. We will use χ_+ and χ_- to indicate if the transition is made with a mutant being selected for birth (χ_+) or resident (χ_-). Hence, for some $C \in V$ and $v \notin C$, $\chi_-(C, C \cup \{v\}) = 0$ and $\chi_+(C \cup \{v\}, C) = 0$. Hence, for all $C \in 2^V$, $\sum_{C' \in 2^V} (\chi_+(C, C') + \chi_-(C, C')) = 1$.
- If the transition function is based on birth-death and computed with some $r > 1$, then we will write it as $\chi_+^{(r)}, \chi_-^{(r)}$ respectively. If computed with $r = 1$, then we write $\chi_+^{(nd)}, \chi_-^{(nd)}$ respectively.

- For some $C \in 2^C$, let $inc(C)$ be the set of all elements $D \in 2^V$ s.t. $|D| \geq |C|$ and $\chi_+(C, D) > 0$.
- For some $C \in 2^C$, let $dec(C)$ be the set of all elements $D \in 2^V$ s.t. $|D| \leq |C|$ and $\chi_-(C, D) > 0$.
- Given set $C \subseteq V$, we will use $F_C^{(r)}$ to denote the probability of fixation given initial set of mutants C where the value r is used to calculate all transition probabilities.

CLAIM 1: If a some time period, the probability distribution over mutant configurations is \mathbf{I} , the fixation probability is $\sum_{C \in 2^V} \mathbf{I}(C) \cdot F_C^{(r)}$.

Clearly, for any time t , $F_C^{(r)} = \lim_{i \rightarrow \infty} \mathbf{Pr}(V^{(i)} | C^{(t)})$. Under the assumption that there exists some time ω s.t. fixation is reached, we have:

$$\begin{aligned} F_C^{(r)} &= \mathbf{Pr}(V^{(\omega)} | C^{(t)}) \\ &= \frac{\mathbf{Pr}(V^{(\omega)} \wedge C^{(t)})}{\mathbf{Pr}(C^{(t)})} \end{aligned}$$

Hence, $F_C^{(r)} \cdot \mathbf{Pr}(C^{(t)}) = \mathbf{Pr}(V^{(\omega)} \wedge C^{(t)})$. The statement then follows by the summation of exhaustive and mutually exclusive events.

CLAIM 2: If a some time period t , the probability distribution over mutant configurations is \mathbf{I} , and the transition functions used to reach the next time step are χ_+, χ_- , then the probability of being in some mutant configuration C at time $t + 1$ is given by $\sum_{D \in 2^V} \mathbf{I}(D) \cdot (\chi_+(D, C) + \chi_-(D, C))$.

Follows directly from the rules of dynamics.

CLAIM 3: If a some time period t , the probability distribution over mutant configurations is \mathbf{I} , mutant fitness r , and the transition functions used to reach the next time step are $\chi_+^{(r)}, \chi_-^{(r)}$, and all subsequent transitions are computed using the same dynamics with neutral drift, then the fixation probability is:

$$\mathcal{P}(I, r) = \sum_{C \in 2^V} \mathbf{I}(C) \cdot \left(\sum_{D \in inc(C)} (\chi_+^{(r)}(C, D) \cdot F_D^{(1)}) + \sum_{D \in dec(C)} (\chi_-^{(r)}(C, D) \cdot F_D^{(1)}) \right)$$

Follows directly from claims 1-2.

CLAIM 4: Under BD-B, BD-D, DB-B, DB-D, or LD dynamics, for some $r \leq r'$, for all $C, D \in 2^V$, we have $\chi_+^{(r)}(C, D) \leq \chi_+^{(r')}(C, D)$ and $\chi_-^{(r)}(C, D) \geq \chi_-^{(r')}(C, D)$.

CLAIM 4a: For some $r \leq r'$, for all $C, D \in 2^V$, we have $\chi_+^{(r)}(C, D) \leq \chi_+^{(r')}(C, D)$.

Let $\{v_j\} = D - C$. For each vertex v_i , $f_i = 1$ if $v_i \notin C$ (a resident) and $f_i = r$ if $v_i \in C$ (a mutant). When $D \equiv C$, the following are all summed over the set $\{v_j \in C \mid \exists v_i \in C \wedge (v_i, v_j) \in E\}$.

- Under BD-B,

$$\chi_+^{(r)}(C, D) = \sum_{\substack{v_i \in C \\ (v_i, v_j) \in E}} \frac{r \cdot w_{ij}}{r \cdot |C| + N - |C|}$$

- Under BD-D,

$$\sum_{\substack{v_i \in C \\ (v_i, v_j) \in E}} \frac{w_{ij}}{N \cdot \sum_{v_q \mid (v_i, v_q) \in E} w_{iq} \cdot f_q^{-1}}$$

- Under DB-B,

$$\chi_+^{(r)}(C, D) = \sum_{\substack{v_i \in C \\ (v_i, v_j) \in E}} \frac{w_{ij} \cdot r}{N \cdot \sum_{v_q \mid (v_q, v_j) \in E} w_{qj} \cdot f_q}$$

- Under DB-D,

$$\chi_+^{(r)}(C, D) = \sum_{\substack{v_i \in C \\ (v_i, v_j) \in E}} \frac{w_{ij}}{\sum_{v_q \in V} f_q^{-1}}$$

- Under LD,

$$\chi_+^{(r)}(C, D) = \sum_{v_i \in C \mid (v_i, v_j) \in E} \frac{w_{ij} \cdot r}{\sum_{v_q, v_\ell \mid (v_q, v_\ell) \in E} w_{q\ell} \cdot f_q}$$

By simple algebraic manipulation, for each of these, when all values other than r are fixed, they increase as r increases.

CLAIM 4b: For some $r \leq r'$, for all $C, D \in 2^V$, we have $\chi_-^{(r)}(C, D) \geq \chi_-^{(r')}(C, D)$. Let $\{v_j\} = C - D$. For each vertex v_i , $f_i = 1$ if $v_i \notin C$ (a resident) and $f_i = r$ if $v_i \in C$ (a mutant). When $D \equiv C$, the following are all summed over the set $\{v_j \in V - C \mid \exists v_i \in V - C \wedge (v_i, v_j) \in E\}$.

- Under BD-B,

$$\chi_+^{(r)}(C, D) = \sum_{\substack{v_i \in V - C \\ (v_i, v_j) \in E}} \frac{w_{ij}}{r \cdot |C| + N - |C|}$$

- Under BD-D,

$$\sum_{\substack{v_i \in V - C \\ (v_i, v_j) \in E}} \frac{w_{ij} \cdot r^{-1}}{N \cdot \sum_{v_q \mid (v_i, v_q) \in E} w_{iq} \cdot f_q^{-1}}$$

- Under DB-B,

$$\chi_+^{(r)}(C, D) = \sum_{\substack{v_i \in V - C \\ (v_i, v_j) \in E}} \frac{w_{ij}}{N \cdot \sum_{v_q \mid (v_q, v_j) \in E} w_{qj} \cdot f_q}$$

- Under DB-D,

$$\chi_+^{(r)}(C, D) = \sum_{\substack{v_i \in V - C \\ (v_i, v_j) \in E}} \frac{w_{ij} \cdot r^{-1}}{\sum_{v_q \in V} f_q^{-1}}$$

- Under LD,

$$\chi_+^{(r)}(C, D) = \sum_{v_i \in V - C \mid (v_i, v_j) \in E} \frac{w_{ij}}{\sum_{v_q, v_\ell \mid (v_q, v_\ell) \in E} w_{q\ell} \cdot f_q}$$

By simple algebraic manipulation, for each of these, when all values other than r are fixed, they decrease as r increases.

CLAIM 5: Given some $C \in 2^V$, for all pairs D, D' where $D \in inc(C)$ and $D' \in dec(C)$, we have $F_D^{(1)} \geq F_{D'}^{(1)}$.

Follows directly from Theorem 5.

CLAIM 6: Given interpretation \mathbf{I} , under BD-B, BD-D, DB-B, DB-D, or LD dynamics, for some $r > 1$, $\mathcal{P}(\mathbf{I}, r) \geq \mathcal{P}(\mathbf{I}, 0)$.

Let us consider some set C from the outermost summation in the computation of $\mathcal{P}(\mathbf{I}, r)$. Suppose, BWOC, there exists some $C \in 2^V$ s.t.:

$$\sum_{D \in inc(C)} (\chi_+^{(r)}(C, D) \cdot F_D^{(1)}) + \sum_{D \in dec(C)} (\chi_-^{(r)}(C, D) \cdot F_D^{(1)}) < \sum_{D \in inc(C)} (\chi_+^{(1)}(C, D) \cdot F_D^{(1)}) + \sum_{D \in dec(C)} (\chi_-^{(1)}(C, D) \cdot F_D^{(1)})$$

This give us:

$$\sum_{D \in inc(C)} (\chi_+^{(r)}(C, D) \cdot F_D^{(1)}) - \sum_{D \in inc(C)} (\chi_+^{(1)}(C, D) \cdot F_D^{(1)}) < \sum_{D \in dec(C)} (\chi_-^{(1)}(C, D) \cdot F_D^{(1)}) - \sum_{D \in dec(C)} (\chi_-^{(r)}(C, D) \cdot F_D^{(1)})$$

Let $F_{sm} = \inf\{F_D^{(1)} | D \in inc(C)\}$ and $F_{lg} = \sup\{F_D^{(1)} | D \in dec(C)\}$, this give us:

$$F_{sm} \sum_{D \in inc(C)} (\chi_+^{(r)}(C, D) - \chi_+^{(1)}(C, D)) < F_{lg} \sum_{D \in dec(C)} (\chi_-^{(1)}(C, D) - \chi_-^{(r)}(C, D))$$

Consider the following:

$$\begin{aligned} \sum_{D \in inc(C)} \chi_+^{(r)}(C, D) + \sum_{D \in dec(C)} \chi_-^{(r)}(C, D) &= \sum_{D \in inc(C)} \chi_+^{(1)}(C, D) + \sum_{D \in dec(C)} \chi_-^{(1)}(C, D) \\ \sum_{D \in inc(C)} \chi_+^{(r)}(C, D) - \sum_{D \in inc(C)} \chi_+^{(1)}(C, D) &= \sum_{D \in dec(C)} \chi_-^{(1)}(C, D) - \sum_{D \in dec(C)} \chi_-^{(r)}(C, D) \end{aligned}$$

Note that by claim 4, both sides of the above equation are positive numbers. Hence, we have $F_{sm} < F_{lg}$, which contradicts claim 5.

PROOF OF THEOREM: Let $\mathcal{P}^{(1)}(\mathbf{I}, r) = \mathcal{P}(\mathbf{I}, r)$ and $\mathcal{P}^{(i+1)}(\mathbf{I}, r) = \mathcal{P}(\mathcal{P}^{(i)}(\mathbf{I}, r), r)$. By claim 6, for any i , $\mathcal{P}^{(i+1)}(\mathbf{I}, r) \geq \mathcal{P}(\mathcal{P}^{(i)}(\mathbf{I}, r), r)$. Consider an interpretation \mathbf{I} that describes the initial probability distribution over mutant configurations. The fixation probability under neutral drift is $\mathcal{P}(\mathbf{I}, 1)$. For some value $r > 1$, the fixation probability is $\lim_{i \rightarrow \infty} \mathcal{P}^{(i)}(\mathbf{I}, r)$. Clearly, $\lim_{i \rightarrow \infty} \mathcal{P}^{(i)}(\mathbf{I}, r) \geq \mathcal{P}(\mathbf{I}, 1)$. Q.E.D.

20. Proof of Theorem 6

$$t_C = \frac{1}{F_C} \sum_{t=1}^{\infty} t \cdot (F_{t|C} - F_{t-1|C})$$

Proof. This proof was first presented in [2]. The mean time to fixation is described as the expected time to fixation given that the process fixates. Let $F_{t|C}$ be the probability that fixation is reached in exactly t time-steps or less. Hence, the probability of reaching fixation in exactly t time steps conditioned on the process reaching fixation is $(F_{t|C} - F_{t-1|C})/F_C$. The remainder of the theorem follows from the definition of an expected value. *Q.E.D.*

21. Proof of Theorem 7

We introduce two pieces of notation t_{fix}, t_{convg} . We define t_{fix} as a time s.t. $t_C = \frac{1}{F_C} \sum_{t=1}^{t_{fix}} t \cdot (F_{t|C} - F_{t-1|C})$. and t_{convg} s.t. $\forall i, \mathbf{Pr}(\mathbf{M}_i^{(t_{convg})}) = F_C$. While in reality, both of these values could be infinite, we note that the relationship $\infty \geq t_{fix} \geq t_{convg}$ holds and that using a lower value for t_{fix} and/or t_{convg} will still ensure we have a lower bound.

$$\frac{1}{F_C} \sum_{t=1}^{\infty} t \cdot (P_{\min,t} - P_{\min,t-1}) \leq \frac{1}{F_C} \sum_{t=1}^{\infty} t \cdot (F_{t|C} - F_{t-1|C})$$

Where $P_{\min,t} = \min_i \mathbf{Pr}(\mathbf{M}_i^{(t)})$.

Proof. First, we have the following.

$$\sum_{t=1}^{t_{convg}} t \cdot (P_{\min,t} - P_{\min,t-1}) \leq \sum_{t=1}^{t_{fix}} t \cdot (F_{t|C} - F_{t-1|C}) \quad (24)$$

For any time t , let $\mathcal{P}_C^{(t)} = F_{t|C} - F_{t-1|C}$ and $\mathbf{Pr}(\Delta \mathbf{M}_{\min}^{(t)}) = P_{\min,t} - P_{\min,t-1}$. As for each time t , we have $P_{\min,t} \geq F_{t|C}$, we can define $\theta_{t'}^{(t)}$ as the portion of $\mathcal{P}_C^{(t)}$ accounted for in $\mathbf{Pr}(\Delta \mathbf{M}_{\min}^{(t')})$. This results in having $\theta_{t'}^{(t)} = 0$ whenever

$t' > t$ or $t > t_{convg}$ as well as the following:

$$\mathcal{P}_C^{(t)} = \sum_{t'=1}^t \theta_{t'}^{(t)} \quad (25)$$

$$\Pr(\Delta \mathbf{M}_{\min}^{(t)}) = \sum_{t'=t}^{t_{fix}} \theta_t^{(t')} \quad (26)$$

$$\sum_{t=1}^{t_{convg}} t \cdot (P_{\min,t} - P_{\min,t-1}) = \sum_{t=1}^{t_{convg}} \sum_{t'=t}^{t_{fix}} t \theta_t^{(t')} \quad (27)$$

$$= \sum_{t=1}^{t_{convg}} \sum_{t'=1}^{t_{fix}} t \theta_t^{(t')} \quad (28)$$

$$= \sum_{t'=1}^{t_{fix}} \sum_{t=1}^{t_{convg}} t \theta_t^{(t')} \quad (29)$$

$$= \sum_{t'=1}^{t_{fix}} \sum_{t=1}^{t'} t \theta_t^{(t')} \quad (30)$$

We also note that the following is true:

$$\sum_{t=1}^{t_{fix}} t \cdot (F_{t|C} - F_{t-1|C}) = \sum_{t'=1}^{t_{fix}} t' \mathcal{P}_C^{(t')} \quad (31)$$

$$= \sum_{t'=1}^{t_{fix}} t' \sum_{t=1}^{t'} \theta_t^{(t')} \quad (32)$$

$$\geq \sum_{t'=1}^{t_{fix}} \sum_{t=1}^{t'} t \theta_t^{(t')} \quad (33)$$

$$= \sum_{t=1}^{t_{convg}} t \cdot (P_{\min,t} - P_{\min,t-1}) \quad (34)$$

Which concludes the proof. *Q.E.D.*

22. Materials and Methods

Except for the experiments dealing with time to fixation/extinction, all algorithms were implemented in Python and run on a 2.33GHz Intel Xeon

CPU. The time-to-fixation experiments were run on a machine equipped with an Intel Core i7 M620 processor running at 2.67 GHz with 4 GB RAM.

All graphs in the experiments were generated using the Python NetworkX package [Hagberg et al.]. Parameters used for the experiments concerning the expected number of mutants were $m=1$ for BA, $p = 0.5$ for ER, and $k = 2$ and $p = 0.5$ for NWS graph generator functions.

We modified Algorithm 1-ACC based on the results on mean time to fixation as follows:

- Before line 14, insert: `t += 1; Sum += t*(min(p)-min(q))`
- Replace line 17 with: `return = Sum/average(p)`, where `average(q)` is the algorithm's best estimate for P_c at termination.